

TOTAL VARIATION ERROR BOUNDS FOR GEOMETRIC APPROXIMATION

EROL PEKÖZ, ADRIAN RÖLLIN AND NATHAN ROSS

*Boston University, National University of Singapore and
University of California, Berkeley*

Abstract

We develop a new formulation of Stein’s method to obtain computable upper bounds on the total variation distance between the geometric distribution and a distribution of interest. Our framework reduces the problem to the construction of a coupling between the original distribution and the “discrete equilibrium” distribution from renewal theory. We illustrate the approach in four nontrivial examples: the geometric sum of independent, non-negative, integer-valued random variables having common mean, the generation size of the critical Galton-Watson process conditioned on non-extinction, the in-degree of a randomly chosen node in the uniform attachment random graph model, and the total degree of both a fixed and randomly chosen node in the preferential attachment random graph model. In the first two examples we obtain error bounds in a metric that is stronger than those available in the literature, and in the final two examples we provide the first explicit bounds.

1 INTRODUCTION

The exponential and geometric distributions are convenient and accurate approximations in a wide variety of complex settings involving rare events, extremes, and waiting times. The difficulty in obtaining explicit error bounds for these approximations beyond an elementary setting is discussed in the preface of Aldous (1989), where the author also points out a lack of such results. Recently, Peköz and Röllin (2011) developed a framework to obtain error bounds for the Kolmogorov and Wasserstein distance metrics between the exponential distribution and a distribution of interest. The main ingredients there are Stein’s method (see Ross and Peköz (2007) for an introduction) along with the equilibrium distribution from renewal theory. Due to

Version from August 25, 2011, 06:50

the flexibility of Stein's method and the close connection between the exponential and geometric distributions, it is natural to attempt to use similar techniques to obtain bounds for the (stronger) total variation distance metric between the geometric distribution and an integer supported distribution. The purpose of this paper is to obtain such bounds having application in both situations where exponential approximation is and is not available.

There are, however, some major complications that arise in trying to carry over approaches for the exponential to the geometric and the stronger total variation metric. To see this, we will first discuss the relationship between our results for approximation by the geometric distribution and the body of literature devoted to approximation by the exponential distribution (see Peköz and Röllin (2011) and references therein). For our purposes, the most pertinent previous efforts focus on determining $d_K(\mathcal{L}(Z), \mathcal{L}(W))$, where Z is an exponential random variable with rate one and W is a mean one random variable; for random variables U and V , we define the Kolmogorov distance by

$$d_K(\mathcal{L}(U), \mathcal{L}(V)) := \sup_x |\mathbb{P}(U \leq x) - \mathbb{P}(V \leq x)|.$$

If W is a non-negative random variable, Z has the exponential distribution with rate one and X has the geometric distribution with parameter p , the triangle inequality and $d_K(\mathcal{L}(Z), \mathcal{L}(pX)) \leq p$ from Peköz and Röllin (2011, Theorem 3.1) give

$$|d_K(\mathcal{L}(Z), \mathcal{L}(W)) - d_K(\mathcal{L}(pX), \mathcal{L}(W))| \leq p. \quad (1.1)$$

Alternatively, in the case that W has the form $Y/\mathbb{E}Y$, where Y is a positive integer-valued random variable, we can compare the distribution of Y to the geometric distribution in the total variation distance, which is a standard measure between the distributions of integer-valued random variables U and V defined by

$$d_{TV}(\mathcal{L}(U), \mathcal{L}(V)) := \sup_{B \subset \mathbb{Z}} |\mathbb{P}(U \in B) - \mathbb{P}(V \in B)|.$$

From this point, it is appropriate to discuss the implications of (1.1) in determining the total variation distance between the distribution of an integer-valued random variable and the geometric distribution. First, note that

$$d_K(\mathcal{L}(U), \mathcal{L}(V)) \leq d_{TV}(\mathcal{L}(U), \mathcal{L}(V)), \quad (1.2)$$

since the supremum on the right hand side is taken over a larger set. Thus, an upper bound on the variation distance between the distribution of two random variables immediately implies the same bound on the Kolmogorov distance. Secondly, there is no useful inequality which is converse to (1.2); for example, if U_n is a random variable which is uniformly distributed on the set of even integers between 1 and $2n$ and V_n is uniformly distributed on the set of odd integers in the same range, we see $d_K(\mathcal{L}(U_n), \mathcal{L}(V_n)) = 1/n$ but $d_{TV}(\mathcal{L}(U_n), \mathcal{L}(V_n)) = 1$. Moreover, there is no canonical method to finesse results from the Kolmogorov distance to the total variation distance. That is, even in the case where approximation by the exponential distribution is fruitful and known results can be applied, it is not clear how to obtain our results below from those existing.

Beside the total variation metric, we will also give bounds on the local metric

$$d_{\text{loc}}(\mathcal{L}(U), \mathcal{L}(V)) := \sup_{m \in \mathbb{Z}} |\mathbb{P}(U = m) - \mathbb{P}(V = m)|.$$

It is clear that d_{loc} will be less than or equal to $\sup_m [\mathbb{P}(U = m) \vee \mathbb{P}(V = m)]$, so that typically better rates need to be obtained in order to provide useful information in this metric.

Our formulation rests on the idea that a positive integer-valued random variable W will be approximately geometrically distributed with parameter $p = 1/\mathbb{E}W$ if $\mathcal{L}(W) \approx \mathcal{L}(W^e)$, where W^e has the (discrete) equilibrium distribution with respect to W defined by

$$\mathbb{P}(W^e \leq k) = \frac{1}{\mathbb{E}W} \sum_{i=1}^k \mathbb{P}(W \geq i), k = 1, 2, \dots \quad (1.3)$$

This distribution arises in discrete-time renewal theory as the time until the next renewal when the process is stationary, and the transformation which maps a distribution to its equilibrium distribution has the geometric distribution with positive support as its unique fixed point. Our main result is an upper bound on the variation distance between the distribution of W and a geometric distribution with parameter $(\mathbb{E}W)^{-1}$, in terms of a coupling between the random variables W^e and W .

This setup is closely related to the exponential approximation formulation of Peköz and Röllin (2011) and also Goldstein (2009), which is also related to the zero-bias transformation of Goldstein and Reinert (1997). As discussed above, a serious difficulty in pushing the results of Peköz and Röllin (2011) through to the stronger total variation metric is that the support of

the distribution to be approximated may not match the support of the geometric distribution well enough. This issue is typical in bounding the total variation distance between integer-valued random variables and can be handled by introducing a term into the bound that quantifies ‘smoothness,’ see, for example Barbour and Čekanavičius (2002); Röllin (2005, 2008). Even with this difficulty, in many of the situations where the ideas of Peköz and Röllin (2011) can be applied, the results here will yield comparable statements in a stronger metric. To illustrate this point, we apply our abstract formulation to obtain new error bounds in two of the examples treated in Peköz and Röllin (2011).

The geometric distribution may also arise as a limiting distribution where exponential approximation is not available. Thus, we will also apply our theory in two examples (discussed in more detail immediately below) that fall into this category. We remark here that these two examples are more naturally suited for geometric approximation so that the technical difficulties discussed above relating to the period of the distribution do not arise.

The first application is a bound on the total variation distance between the geometric distribution and the sum of a geometrically distributed number of independent, non-negative, integer-valued random variables with common mean. The distribution of such geometric convolutions have been considered in many places in the literature in the setting of exponential approximation and convergence; the book-length treatment is given in Kalashnikov (1997). The second application is a variation on the classical theorem of Yaglom (1947) describing the asymptotic behavior of the generation size of a critical Galton-Watson process conditioned on non-extinction. This theorem has a large literature of extensions and embellishments (see Lalley and Zheng (in press) for example). Peköz and Röllin (2011) obtained a rate of convergence for the Kolmogorov distance between the generation size of a critical Galton-Watson process conditioned on non-extinction and the exponential distribution. Here we obtain an analogous bound for the geometric distribution in total variation distance. The third application is to the in-degree of a randomly chosen node in the uniform attachment random graph discussed in Bollobas et al. (2001), and the final application is to the total degree of both a fixed and a randomly chosen node in the preferential attachment random graph discussed in Bollobas et al. (2001). In contrast to our first two examples, these examples do not derive from an exponential approximation result.

Finally, we mention that there are other formulations of geometric approximation using Stein’s method. For example, Peköz (1996) and Barbour and Grübel (1995) use the intuition that a positive, integer-valued

random variable W approximately has a geometric distribution with parameter $p = \mathbb{P}(W = 1)$ if

$$\mathcal{L}(W) \approx \mathcal{L}(W - 1 | W > 1).$$

Other approaches can be found in Phillips and Weinberg (2000) and Daly (2010).

The organization of this article is as follows. In Section 2 we present our main theorems, and Sections 3, 4, 5, and 6 respectively contain applications to geometric sums, the critical Galton Watson process conditioned on non-extinction, the uniform attachment random graph model, and the preferential attachment random graph model.

2 MAIN RESULTS

A typical issue when discussing the geometric distribution is whether to have the support begin at zero or one. Denote by $\text{Ge}(p)$ the geometric distribution with positive support; that is $\mathcal{L}(Z) = \text{Ge}(p)$ if $\mathbb{P}(Z = k) = (1 - p)^{k-1}p$ for positive integers k . Alternatively, denote by $\text{Ge}^0(p)$ the geometric distribution $\text{Ge}(p)$ shifted by minus one, that is “starting at 0.” Since $\mathcal{L}(Z) = \text{Ge}(p)$, implies $\mathcal{L}(Z - 1) = \text{Ge}^0(p)$, it is typical that results for one of $\text{Ge}(p)$ or $\text{Ge}^0(p)$ easily pass to the other. Unfortunately, our methods do not appear to trivially transfer between these two distributions, so we are forced to develop our theory for both cases in parallel.

First, we give an alternate definition of the equilibrium distribution that we will use in the proof of our main result.

Definition 2.1. Let X be a positive, integer-valued random variable with finite mean. We say that an integer-valued random variable X^e has the *discrete equilibrium distribution w.r.t. X* if for all bounded f and $\nabla f(x) = f(x) - f(x - 1)$ we have

$$\mathbb{E}f(X) - f(0) = \mathbb{E}X \mathbb{E}\nabla f(X^e). \quad (2.1)$$

Remark 2.2. To see how (2.1) is equivalent to (1.3), note that we have

$$\mathbb{E}f(X) - f(0) = \mathbb{E} \sum_{i=1}^X \nabla f(i) = \sum_{i=1}^{\infty} \nabla f(i) \mathbb{P}(X \geq i) = \mathbb{E}X \mathbb{E}\nabla f(X^e).$$

In order to handle non-negative random variables, we also introduce a variation of Definition 2.1.

Definition 2.3. If X is a non-negative integer-valued random variable with $\mathbb{P}(X = 0) > 0$, we say that an integer-valued random variable X^{e_0} has the *discrete equilibrium distribution w.r.t. X* if for all bounded f and with $\Delta f(x) = f(x+1) - f(x)$ we have

$$\mathbb{E}f(X) - f(0) = \mathbb{E}X \mathbb{E}\Delta f(X^{e_0}).$$

Note that we are defining the term “discrete equilibrium distribution” in both of the previous definitions, but this should not cause confusion as the support of the base distribution dictates the meaning of the terminology.

As a final bit of notation before the statement of our main results, for a function g with domain \mathbb{Z} , let $\|g\| = \sup_{k \in \mathbb{Z}} |g(k)|$, and for any integer-valued random variable W and any σ -algebra \mathcal{F} , define the conditional smoothness

$$\begin{aligned} S_1(W|\mathcal{F}) &= \sup_{\|g\| \leq 1} |\mathbb{E}\{\Delta g(W)|\mathcal{F}\}| \\ &= 2 d_{\text{TV}}(\mathcal{L}(W+1|\mathcal{F}), \mathcal{L}(W|\mathcal{F})), \end{aligned} \tag{2.2}$$

and the second order conditional smoothness

$$S_2(W|\mathcal{F}) = \sup_{\|g\| \leq 1} |\mathbb{E}\{\Delta^2 g(W)|\mathcal{F}\}|,$$

where $\Delta^2 g(k) = \Delta g(k+1) - \Delta g(k)$. In order to simplify the presentation of the main theorems, we let $d_1 = d_{\text{TV}}$ and $d_2 = d_{\text{loc}}$.

Theorem 2.1. *Let W be a positive integer-valued random variable with $\mathbb{E}W = 1/p$ for some $0 < p \leq 1$ and let W^e have the discrete equilibrium distribution w.r.t. W . Then with $D = W - W^e$, any σ -algebra $\mathcal{F} \supseteq \sigma(D)$ and event $A \in \mathcal{F}$ we have*

$$d_l(\mathcal{L}(W), \text{Ge}(p)) \leq \mathbb{E}\{|D|S_l(W|\mathcal{F})\mathbf{I}_A\} + 2\mathbb{P}(A^c), \tag{2.3}$$

for $l = 1, 2$, and

$$d_{\text{TV}}(\mathcal{L}(W^e), \text{Ge}(p)) \leq p\mathbb{E}|D|, \tag{2.4}$$

$$d_{\text{loc}}(\mathcal{L}(W^e), \text{Ge}(p)) \leq p\mathbb{E}\{|D|S_1(W|\mathcal{F})\}; \tag{2.5}$$

on the RHS of (2.3) and (2.5), $S_l(W|\mathcal{F})$ can be replaced by $S_l(W^e|\mathcal{F})$.

Theorem 2.2. *Let W be a non-negative integer-valued random variable with $\mathbb{P}(W = 0) > 0$, $\mathbb{E}W = (1-p)/p$ for some $0 < p \leq 1$ and let W^{e_0} have the*

discrete equilibrium distribution w.r.t. W . Then with $D = W - W^{e_0}$, any σ -algebra $\mathcal{F} \supseteq \sigma(D)$ and event $A \in \mathcal{F}$ we have

$$d_l(\mathcal{L}(W), \text{Ge}^0(p)) \leq (1-p)\mathbb{E}\{|D|S_l(W|\mathcal{F})\mathbf{I}_A\} + 2(1-p)\mathbb{P}(A^c) \quad (2.6)$$

for $l = 1, 2$, and

$$d_{\text{TV}}(\mathcal{L}(W^{e_0}), \text{Ge}^0(p)) \leq p\mathbb{E}|D|, \quad (2.7)$$

$$d_{\text{loc}}(\mathcal{L}(W^{e_0}), \text{Ge}^0(p)) \leq p\mathbb{E}\{|D|S_1(W|\mathcal{F})\}, \quad (2.8)$$

on the RHS of (2.6) and (2.8), $S_l(W|\mathcal{F})$ can be replaced by $S_l(W^{e_0}|\mathcal{F})$.

Before we prove Theorems 2.1 and 2.2, we make a few remarks related to these results.

Remark 2.4. It is easy to see that the a random variable W with law equal to $\text{Ge}(p)$ has the property that $\mathcal{L}(W) = \mathcal{L}(W^e)$, so that W^e can be taken to be W and the theorem yields the correct error term in this case. The analogous statement is true for $\text{Ge}^0(p)$ and W^{e_0} .

Remark 2.5. By choosing $A = \{W = W^{e_0}\}$ in (2.6) with $l = 1$, we find

$$d_{\text{TV}}(\mathcal{L}(W), \text{Ge}^0(p)) \leq 2(1-p)\mathbb{P}(W \neq W^{e_0}), \quad (2.9)$$

and an analogous corollary holds for Theorem 2.1.

In order to use the theorem we need to be able to construct random variables with the discrete equilibrium distribution. The next proposition provides such a construction for a non-negative integer valued random variable W . We say W^s has the size-bias distribution of W , if

$$\mathbb{E}\{Wf(W)\} = \mathbb{E}W\mathbb{E}f(W^s)$$

for all f for which the expectation exist.

Proposition 2.3. *Let W be an integer-valued random variable and let W^s have the size-bias distribution of W .*

1. *If $W > 0$ and we define the random variable W^e such that conditional on W^s , W^e has the uniform distribution on the integers $\{1, 2, \dots, W^s\}$, then W^e has the discrete equilibrium distribution w.r.t. W .*
2. *If $W \geq 0$ with $\mathbb{P}(W = 0) > 0$ and we define the random variable W^{e_0} such that conditional on W^s , W^{e_0} has the uniform distribution on the integers $\{0, 1, \dots, W^s - 1\}$, then W^{e_0} has the discrete equilibrium distribution w.r.t. W .*

Proof. For any bounded f we have

$$\mathbb{E}f(W) - f(0) = \mathbb{E} \sum_{i=1}^W \nabla f(i) = \mathbb{E}W \mathbb{E} \left\{ \frac{1}{W^s} \sum_{i=1}^{W^s} \nabla f(i) \right\} = \mathbb{E}W \mathbb{E} \nabla f(W^e),$$

which implies Item 1. The second item is proved analogously. \square

As mentioned in the introduction, there can be considerable technical difficulty in ensuring the support of the distribution to be approximated is smooth. In Theorems 2.1 and 2.2 this issue is accounted for in the term $S_1(W|\mathcal{F})$. Typically, our strategy to bound this term will be to write W (or W^e) as a sum of terms which are independent given \mathcal{F} and then apply the following lemma from Mattner and Roos (2007, Corollary 1.6).

Lemma 2.4. (*Mattner and Roos (2007, Corollary 1.6)*) *If X_1, \dots, X_n are independent integer-valued random variables and*

$$u_i = 1 - d_{\text{TV}}(\mathcal{L}(X_i), \mathcal{L}(X_i + 1))$$

then

$$d_{\text{TV}}\left(\mathcal{L}\left(\sum_{i=1}^n X_i\right), \mathcal{L}\left(1 + \sum_{i=1}^n X_i\right)\right) \leq \sqrt{\frac{2}{\pi}} \left(\frac{1}{4} + \sum_{i=1}^n u_i\right)^{-1/2}.$$

Remark 2.6. It may seem surprising that the bound (2.4) gains an extra factor of p , which is considered small if the geometric distribution is approximately exponential. To explain this phenomenon, Proposition 2.3 indicates that W^e is already ‘smooth’ in the sense that there are no gaps in its support. More precisely, the analog of the term in (2.3) accounting for the period of the support of W^e is automatically small. Heuristically, note that from (2.1) we have that

$$|\mathbb{E} \nabla f(W^e)| \leq \frac{2\|f\|}{\mathbb{E}W}$$

which implies that $d_{\text{TV}}(\mathcal{L}(W^e), \mathcal{L}(W^e + 1))$ is of order $(\mathbb{E}W)^{-1}$, regardless of the distribution of W .

Before we present the proof of Theorems 2.1 and 2.2, we must first develop the Stein’s method machinery we will need. As in Peköz (1996), for any subset B of the integers and any $p = 1 - q$ we construct the function $f = f_{B,p}$ defined by $f(0) = 0$ and for $k \geq 1$,

$$qf(k) - f(k-1) = I_{k \in B} - \text{Ge}(p)\{B\}, \quad (2.10)$$

where $\text{Ge}(p)\{B\} = \sum_{i \in B} (1-p)^{i-1} p$ is the chance that a positive geometric random variable with parameter p takes a value in the set B . It can be easily verified that the solution of (2.10) is given by

$$f(k) = \sum_{i \in B} q^{i-1} - \sum_{i \in B, i \geq k+1} q^{i-k-1}. \quad (2.11)$$

Equivalently, for $k \geq 0$,

$$qf(k+1) - f(k) = I_{k \in B-1} - \text{Ge}^0(p)\{B-1\},$$

where we define $\text{Ge}^0(p)\{B\}$ analogous to $\text{Ge}(p)\{B\}$.

Peköz (1996) and Daly (2008) study properties of these solutions, but we next need the following additional lemma to obtain our main result.

Lemma 2.5. *For f as above, we have*

$$\sup_{k \geq 1} |\nabla f(k)| = \sup_{k \geq 0} |\Delta f(k)| \leq 1. \quad (2.12)$$

If, in addition, $B = \{m\}$ for some $m \in \mathbb{Z}$, then

$$\sup_{k \geq 0} |f(k)| \leq 1. \quad (2.13)$$

Proof. To show (2.12), note that

$$\begin{aligned} \nabla f(k) &= \sum_{i \in B, i \geq k} q^{i-k} - \sum_{i \in B, i \geq k+1} q^{i-k-1} \\ &= I_{k \in B} + \sum_{i \in B, i \geq k+1} (q^{i-k} - q^{i-k-1}) \\ &= I_{k \in B} - p \sum_{i \in B, i \geq k+1} q^{i-k-1}, \end{aligned}$$

thus $-1 \leq \nabla f(k) \leq 1$. If now $B = \{m\}$, (2.13) is immediate from (2.11). \square

We are now ready to present the proof of our main results.

Proof of Theorem 2.1. Given any positive integer-valued random variable W with $\mathbb{E}W = 1/p$ and $D = W - W^e$ we have, using (2.10), Definition 2.1,

and Lemma 2.5 in the two inequalities,

$$\begin{aligned}
& \mathbb{P}(W \in B) - \text{Ge}(p)\{B\} \\
&= \mathbb{E}\{qf(W) - f(W-1)\} \\
&= \mathbb{E}\{\nabla f(W) - pf(W)\} \\
&= \mathbb{E}\{\nabla f(W) - \nabla f(W^e)\} \\
&\leq \mathbb{E}\{I_A(\nabla f(W) - \nabla f(W^e))\} + 2\mathbb{P}(A^c) \\
&= \mathbb{E}\left\{I_A I[D > 0] \sum_{i=0}^{D-1} \mathbb{E}(\nabla f(W^e + i + 1) - \nabla f(W^e + i) | \mathcal{F})\right\} \\
&\quad + \mathbb{E}\left\{I_A I[D < 0] \sum_{i=0}^{-D-1} \mathbb{E}(\nabla f(W^e - i - 1) - \nabla f(W^e - i) | \mathcal{F})\right\} + 2\mathbb{P}(A^c) \\
&\leq \mathbb{E}\{|D|S_1(W^e | \mathcal{F})I_A\} + 2\mathbb{P}(A^c),
\end{aligned}$$

which is (2.3) for $l = 1$; analogously, one can obtain (2.3) with $S_1(W | \mathcal{F})$ in place of $S_1(W^e | \mathcal{F})$ on the RHS. In the case of $B = \{m\}$ we can make use of (2.13) to obtain

$$|\mathbb{P}(W = m) - \text{Ge}(p)\{m\}| \leq \mathbb{E}\{|D|S_2(W^e | \mathcal{F})I_A\} + 2\mathbb{P}(A^c),$$

instead, which proves (2.3) for $l = 2$. For (2.4), we have that

$$\begin{aligned}
\mathbb{P}(W^e \in B) - \text{Ge}(p)\{B\} &= \mathbb{E}\{qf(W^e) - f(W^e - 1)\} \\
&= \mathbb{E}\{\nabla f(W^e) - pf(W^e)\} = p\mathbb{E}\{f(W) - f(W^e)\} \leq p\mathbb{E}|D|,
\end{aligned}$$

where the last line follows by writing $f(W) - f(W^e)$ as a telescoping sum of $|D|$ terms no greater than $\|\nabla f\|$, which can be bounded using (2.12); (2.5) is straightforward using (2.13) and (2.2). \square

Proof of Theorem 2.2. Let W be a non-negative integer-valued random variable with $\mathbb{E}W = (1-p)/p$ and W^{e_0} as in the theorem. If we define I to be independent of all else and such that $\mathbb{P}(I = 1) = 1 - \mathbb{P}(I = 0) = p$, then a short calculation shows that the variable defined by

$$[(W+1)^e | I = 1] = W+1 \quad \text{and} \quad [(W+1)^e | I = 0] = W^{e_0} + 1 \quad (2.14)$$

has the positive discrete equilibrium transform with respect to $W+1$. Equation (2.6) now follows after noting that

$$d_l(\mathcal{L}(W), \text{Ge}^0(p)) = d_l(\mathcal{L}(W+1), \text{Ge}(p)),$$

and then applying the following stronger variation of (2.3) which is easily read from the proof of Theorem 2.1:

$$d_l(\mathcal{L}(W+1), \text{Ge}(p)) \leq \mathbb{E}\{|W+1 - (W+1)^e| S_l(W|\mathcal{F}) \mathbf{I}_A\} \\ + 2\mathbb{P}((W+1)^e \neq W+1) \mathbb{P}(A^c).$$

Equations (2.7) and (2.8) can be proved in a manner similar to their analogs in Theorem 2.1. □

Remark 2.7. As the proof of Theorem 2.2 shows, (2.14) can be used to finesse results using the coupling $((W+1), (W+1)^e)$ for W non-negative to results using the coupling (W, W^{e_0}) . However, it is not clear how to implement this approach in application; for example in Section 3 below.

3 APPLICATIONS TO GEOMETRIC SUMS

In this section we apply the results above to a sum of the geometric number of independent but not necessarily identically distributed random variables. As in our theory above, we will have separate results for the two cases where the sum is strictly positive and the case where it can take on the value zero with positive probability. We reiterate that although there are a variety of exponential approximation results in the literature for this example, there do not appear to be bounds available for the analogous geometric approximation in the total variation metric.

Theorem 3.1. *Let X_1, X_2, \dots be a sequence of independent, square integrable, positive, and integer-valued random variables, such that for some $u > 0$ we have, for all $i \geq 1$, $\mathbb{E}X_i = \mu$ and $u \leq 1 - d_{\text{TV}}(\mathcal{L}(X_i), \mathcal{L}(X_i+1))$. Let $\mathcal{L}(N) = \text{Ge}(a)$ for some $0 < a \leq 1$ and $W = \sum_{i=1}^N X_i$. Then with $p = 1 - q = a/\mu$, we have*

$$d_l(\mathcal{L}(W), \text{Ge}(p)) \leq C_l \sup_{i \geq 1} \mathbb{E}|X_i - X_i^e| \leq C_l(\mu_2/2 + \frac{1}{2} + \mu) \quad (3.1)$$

for $l = 1, 2$, where $\mu_2 := \sup_i \mathbb{E}X_i^2$ and

$$C_1 = \min\left\{1, a\left[1 + \left(-\frac{2}{u \log(1-a)}\right)^{1/2}\right]\right\}, \\ C_2 = \min\left\{1, a\left[1 - \frac{6 \log(a)}{\pi u}\right]\right\}.$$

Theorem 3.2. Let X_1, X_2, \dots be a sequence of independent, square integrable, non-negative and integer-valued random variables, such that for some $u > 0$ we have, for all $i \geq 1$, $\mathbb{E}X_i = \mu$ and $u \leq 1 - d_{\text{TV}}(\mathcal{L}(X_i), \mathcal{L}(X_i + 1))$. Let $\mathcal{L}(M) = \text{Ge}^0(a)$ for some $0 < a \leq 1$ and $W = \sum_{i=1}^M X_i$. Then with $p = 1 - q = a/(a + \mu(1 - a))$, we have

$$d_l(\mathcal{L}(W), \text{Ge}^0(p)) \leq C_l \sup_{i \geq 1} \mathbb{E}X_i^{e_0} \leq C_l (\mu_2/(2\mu) - \frac{1}{2}), \quad (3.2)$$

for $l = 1, 2$, where $\mu_2 := \sup_i \mathbb{E}X_i^2$ and the C_l are as in Theorem 3.1.

Before proving Theorems 3.1 and 3.2, we make a few remarks.

Remark 3.1. The first inequality in (3.1) yields the correct bound of zero when X_i is geometric, as in this case we would have $X_i = X_i^e$; see Remark 2.4 following Theorem 2.1. Similarly, in the case where the X_i have a Bernoulli distribution with expectation μ , we have that $X^{e_0} = 0$ so that the left hand side of (3.2) is zero. That is, if $M \sim \text{Ge}^0(a)$ and conditional on M , W has the binomial distribution with parameters M and μ for some $0 \leq \mu \leq 1$, then $W \sim \text{Ge}^0(a/(a + \mu(1 - a)))$.

Remark 3.2. The local metric result in Theorem 3.1 above will not be useful in the regime of exponential distribution convergence where we only assume a is considered small, since the probabilities being approximated are of smaller order (linear in a) than the error bound ($a \log(1/a)$). However, in the case that the X_i are close to geometric, the local metric result may yield useful information.

Remark 3.3. In the case where X_i are i.i.d. but not necessarily integer valued and $0 < a \leq \frac{1}{2}$, Brown (1990, Theorem 2.1) obtains the exponential approximation result

$$d_K(\mathcal{L}(W), \text{Exp}(1/p)) \leq \frac{a\mu_2}{\mu} \quad (3.3)$$

for the weaker Kolmogorov metric. To compare (3.3) with (3.1) for small a , we observe that the bound (3.1) is linear in a whereas (3.3), within a constant factor, behaves like $a(-\log(1 - a))^{-1/2} \sim \sqrt{a}$. Therefore the bound (3.3) is better, but (3.1) applies to non-i.i.d. random variables (albeit having identical means) and to the stronger total variation metric.

Proof of Theorem 3.1. First, let us prove that $W^e := \sum_{i=1}^{N-1} X_i + X_N^e$ has the discrete equilibrium distribution w.r.t. W , where, for each $i \geq 1$, X_i^e is a

random variable having the equilibrium distribution w.r.t. X_i , independent of all else. Note first that we have for bounded f and every m ,

$$\mu \mathbb{E} \nabla f \left(\sum_{i=1}^{m-1} X_i + X_m^e \right) = \mathbb{E} \left[f \left(\sum_{i=1}^m X_i \right) - f \left(\sum_{i=1}^{m-1} X_i \right) \right].$$

Note also that since N is geometric, for any bounded function g with $g(0) = 0$ we have

$$\mathbb{E} \{ g(N) - g(N-1) \} = a \mathbb{E} g(N).$$

We now assume that $f(0) = 0$. Hence, using the above two facts and independence between N and the sequence X_1, X_2, \dots , we have

$$\begin{aligned} \mathbb{E} W \mathbb{E} \nabla f(W^e) &= \frac{\mu}{a} \mathbb{E} \nabla f \left(\sum_{i=1}^{N-1} X_i + X_N^e \right) \\ &= \frac{1}{a} \mathbb{E} \left[f \left(\sum_{i=1}^N X_i \right) - f \left(\sum_{i=1}^{N-1} X_i \right) \right] \\ &= \mathbb{E} f \left(\sum_{i=1}^N X_i \right) = \mathbb{E} f(W). \end{aligned}$$

Now,

$$D = W - W^e = X_N - X_N^e,$$

and setting $\mathcal{F} = \sigma(N, X_N^e, X_N)$, we have

$$\begin{aligned} S_1(W^e | \mathcal{F}) &= S_1 \left(\sum_{i=1}^{N-1} X_i \mid \mathcal{F} \right) \leq 1 \wedge \left(\frac{2}{\pi(0.25 + (N-1)u)} \right)^{1/2} \\ &\leq 1 \wedge \left(\frac{2}{\pi(N-1)u} \right)^{1/2}, \end{aligned} \quad (3.4)$$

where we have used Lemma 2.4 and the fact that $S_1(W^e | \mathcal{F})$ is almost surely bounded by one. We now have

$$\mathbb{E}[|D| S_1(W^e | \mathcal{F})] \leq \mathbb{E} \left[\left(1 \wedge \left(\frac{2}{\pi(N-1)u} \right)^{1/2} \right) \mathbb{E}^N |X_N - X_N^e| \right]. \quad (3.5)$$

From here, we can obtain the first inequality in (3.1) by applying Theorem 2.1 after noting that

$$\mathbb{E}^N |X_N - X_N^e| \leq \sup_{i \geq 1} \mathbb{E} |X_i - X_i^e|,$$

and

$$\begin{aligned}
\mathbb{E}\left(1 \wedge \left(\frac{2}{\pi(N-1)u}\right)^{1/2}\right) &\leq 1 \wedge \left(a + \left(\frac{2}{\pi u}\right)^{1/2} \sum_{i \geq 1} \frac{a(1-a)^i}{i^{1/2}}\right) \\
&\leq 1 \wedge \left(a + a \left(\frac{2}{\pi u}\right)^{1/2} \left(-\frac{\pi}{\log(1-a)}\right)^{1/2}\right) \\
&= 1 \wedge \left(a \left[1 + \left(-\frac{2}{\log(1-a)u}\right)^{1/2}\right]\right),
\end{aligned}$$

where we have used

$$\sum_{i \geq 1} \frac{a(1-a)^{i-1}}{i^{1/2}} \leq \frac{a}{1-a} \int_0^\infty \frac{(1-a)^x}{x^{1/2}} dx = \frac{a}{1-a} \left(-\frac{\pi}{\log(1-a)}\right)^{1/2}.$$

The second inequality in (3.1) follows from Theorem 2.1 and the fact (from the definition of the transformation X^e) that $\mathbb{E}^N |X_N - X_N^e| \leq \mu_2/2 + \frac{1}{2} + \mu$.

To obtain the local limit result, note that, if $V = X + Y$ is the sum of two independent random variables, then $S_2(V) \leq S_1(X)S_1(Y)$. Hence,

$$S_2(W^e|\mathcal{F}) \leq 1 \wedge \frac{2}{\pi(0.25 + (N/2 - 1)_+ u)} \leq 1 \wedge \frac{6}{\pi(N-1)u}.$$

From here we have

$$\mathbb{E}S_2(W^e|\mathcal{F}) \leq 1 \wedge \left(a + \frac{6}{\pi u} \sum_{i \geq 1} \frac{a(1-a)^i}{i}\right) = 1 \wedge \left(a - \frac{6a \log(a)}{\pi u}\right)$$

□

Proof of Theorem 3.2. It is straightforward to check that $W^{e_0} := \sum_{i=1}^M X_i + X_{M+1}^{e_0}$ has the equilibrium distribution with respect to W . Now,

$$D = W - W^{e_0} = -X_{M+1}^{e_0},$$

and setting $\mathcal{F} = \sigma(M, X_{M+1}^{e_0})$, we have

$$S_1(W^{e_0}|\mathcal{F}) = S_1\left(\sum_{i=1}^M X_i \middle| \mathcal{F}\right)$$

which can be bounded above by (3.4) as in the proof of Theorem 3.1. The remainder of the proof follows closely to that of Theorem 3.1. For example, the expression analogous to (3.5) is

$$\mathbb{E}[|D|S_1(W^{e_0}|\mathcal{F})] \leq \mathbb{E} \left[\min \left\{ 1, \frac{\sqrt{2}}{(\pi Mu)^{1/2}} \right\} \mathbb{E}^M X_M^{e_0} \right],$$

and the definition of the transform X^{e_0} implies that

$$\mathbb{E} X_i^{e_0} = \frac{\mathbb{E} X_i^2}{2\mu} - \frac{1}{2}.$$

□

4 APPLICATION TO THE CRITICAL GALTON-WATSON BRANCHING PROCESS

Let $Z_0 = 1, Z_1, Z_2, \dots$ be a Galton-Watson branching process with offspring distribution $\mathcal{L}(Z_1)$. A theorem due to Yaglom (1947) states that, if $\mathbb{E} Z_1 = 1$ and $\text{Var } Z_1 = \sigma^2 < \infty$, then $\mathcal{L}(n^{-1} Z_n | Z_n > 0)$ converges to an exponential distribution with mean $\sigma^2/2$. The recent article Peköz and Röllin (2011) is the first to give an explicit bound on the rate of convergence for this asymptotic result. Using ideas from there, we give a convergence rate for the total variation error of a geometric approximation to Z_n under finite third moment of the offspring distribution and the natural periodicity requirement that

$$d_{\text{TV}}(\mathcal{L}(Z_1), \mathcal{L}(Z_1 + 1)) < 1. \quad (4.1)$$

This type of smoothness condition is typical in the context of Stein's method for approximation by a discrete distribution; see e.g. Barbour and Čekanavičius (2002) and Röllin (2008).

For the proof of the following theorem, we make use the of construction of Lyons, Pemantle, and Peres (1995); we refer to that article for more details on the construction and only present what is needed for our purpose.

Theorem 4.1. *For a critical Galton-Watson branching process with offspring distribution $\mathcal{L}(Z_1)$ such that $\mathbb{E} Z_1^3 < \infty$ and (4.1) holds, we have*

$$d_{\text{TV}}(\mathcal{L}(Z_n | Z_n > 0), \text{Ge}(\frac{2}{\sigma^2 n})) \leq \frac{C \log n}{n^{1/4}} \quad (4.2)$$

for some constant C which is independent of n .

Remark 4.1. Peköz and Röllin (2011, Theorem 4.1) gives the result

$$d_K(\mathcal{L}(2Z_n/(\sigma^2 n)|Z_n > 0), \text{Exp}(1)) \leq C \left(\frac{\log n}{n} \right)^{1/2} \quad (4.3)$$

without Condition (4.1) for the weaker Kolmogorov metric. It can be seen that the bound in (4.2) is not as good as the bound in (4.3) for large n , but (4.2) applies to the stronger total variation metric.

Proof of Theorem 4.1. First we construct a size-biased branching tree as in Lyons et al. (1995). We assume that this tree is labeled and ordered, in the sense that, if w and v are vertices in the tree from the same generation and w is to the left of v , then the offspring of w is to the left of the offspring of v . Start in generation 0 with one vertex v_0 and let it have a number of offspring distributed according to the size-bias distribution of $\mathcal{L}(Z_1)$. Pick one of the offspring of v_0 uniformly at random and call it v_1 . To each of the siblings of v_1 attach an independent Galton-Watson branching process with offspring distribution $\mathcal{L}(Z_1)$. For v_1 proceed as for v_0 , i.e., give it a size-biased number of offspring, pick one uniformly at random, call it v_2 , attach independent Galton-Watson branching process to the siblings of v_2 and so on. It is clear that this will always give an infinite tree as the “spine” v_0, v_1, v_2, \dots is an infinite sequence where v_i is an individual (or particle) in generation i .

We now need some notation. Denote by S_n the total number of particles in generation n . Denote by L_n and R_n the number of particles to the left (excluding v_n) and to the right (including v_n), of vertex v_n , so that $S_n = L_n + R_n$. We can describe these particles in more detail, according to the generation at which they split off from the spine. Denote by $S_{n,j}$ the number of particles in generation n that stem from any of the siblings of v_j (but not v_j itself). Clearly, $S_n = 1 + \sum_{j=1}^n S_{n,j}$, where the summands are independent. Likewise, let $L_{n,j}$ and $R_{n,j}$ be the number of particles in generation n that stem from the siblings to the left and right of v_j (note that $L_{n,n}$ and $R_{n,n}$ are just the number of siblings of v_n to the left and to the right, respectively). We have the relations $L_n = \sum_{j=1}^n L_{n,j}$ and $R_n = 1 + \sum_{j=1}^n R_{n,j}$. Note that, for fixed j , $L_{n,j}$ and $R_{n,j}$ are in general not independent, as they are linked through the offspring size of v_{j-1} .

Let now $R'_{n,j}$ be independent random variables such that

$$\mathcal{L}(R'_{n,j}) = \mathcal{L}(R_{n,j}|L_{n,j} = 0).$$

and, with $A_{n,j} = \{L_{n,j} = 0\}$, define

$$R_{n,j}^* = R_{n,j}I_{A_{n,j}} + R'_{n,j}I_{A_{n,j}^c} = R_{n,j} + (R'_{n,j} - R_{n,j})I_{A_{n,j}^c}.$$

Define also $R_n^* = 1 + \sum_{j=1}^n R_{n,j}^*$. Let us collect a few facts from Peköz and Röllin (2011) which we will then use to give the proof of the theorem (here and in the rest of the proof, C shall denote a constant which is independent of n , but may depend on $\mathcal{L}(Z_1)$ and may also be different from formula to formula):

- (i) $\mathcal{L}(R_n^*) = \mathcal{L}(Z_n | Z_n > 0)$;
- (ii) S_n has the size-biased distribution of Z_n ,
and v_n is equally likely to be any of the S_n particles;
- (iii) $\mathbb{E}\{R'_{n,j} I_{A_{n,j}^c}\} \leq \sigma^2 \mathbb{P}[A_{n,j}^c]$;
- (iv) $\mathbb{E}\{R_{n,j} I_{A_{n,j}^c}\} \leq \gamma \mathbb{P}[A_{n,j}^c]$, and $\mathbb{E}\{R_{n-1,j} I_{A_{n,j}^c}\} \leq \gamma \mathbb{P}[A_{n,j}^c]$,
where $\gamma = \mathbb{E}Z_1^3$;
- (v) $\mathbb{P}[A_{n,j}^c] \leq \sigma^2 \mathbb{P}[Z_{n-j} > 0] \leq C/(n-j+1)$ for some $C > 0$.

In light of (i) and (ii) (and then using the construction in Proposition 2.3 to

see that R_n has the discrete equilibrium distribution w.r.t $\mathcal{L}(R_n^*)$) we can let $W = R_n^*$, $W^e = R_n$ and

$$D = R_n^* - R_n = \sum_{j=1}^n (R'_{n,j} - R_{n,j}) I_{A_{n,j}^c}.$$

Also let

$$N = \sum_{j=1}^{n-1} R_{n-1,j} I_{A_{n,j}^c} \quad \text{and} \quad M = \sum_{j=1}^n R_{n,j} I_{A_{n,j}^c}$$

and note that (iii)–(v) give

$$\mathbb{E}|D| \leq C \log n \quad \text{and} \quad \mathbb{E}N \leq C \log n. \quad (4.4)$$

Next with $\mathcal{F} = \sigma(N, D, R_{n-1}, M, R_{n,n} \mathbf{I}_{A_{n,n}})$ and, letting $Z_1^i, i = 1, 2, \dots$, be i.i.d. copies of Z_1 , we have

$$\mathcal{L}(R_n - M - R_{n,n} \mathbf{I}_{A_{n,n}} - 1 | \mathcal{F}) = \mathcal{L}\left(\sum_{i=1}^{R_{n-1}-N} Z_1^i \mid R_{n-1}, N\right),$$

which follows since $R_n - M = 1 + \sum_{i=1}^n R_{n,j} \mathbf{I}_{A_{n,j}}$ and the particles counted by $R_{n-1} - N$ will be parents of the particles counted by $R_n - M - 1 + R_{n,n} \mathbf{I}_{A_{n,n}}$.

Then we use Lemma 2.4 to obtain

$$\begin{aligned} S_1(W^e|\mathcal{F}) &= S_1(R_n - M - R_{n,n}\mathbf{I}_{A_{n,n}} - 1|\mathcal{F}) \\ &\leq \frac{0.8}{(0.25 + (R_{n-1} - N)u)^{1/2}}. \end{aligned} \quad (4.5)$$

As a direct corollary of (2.4), for any bounded function f we have

$$\mathbb{E}f(W^e) \leq \mathbb{E}f(X_p) + p\|f\|\mathbb{E}|W^e - W|, \quad (4.6)$$

where $X_p \sim \text{Ge}(p)$. Fix $q = 1/\mathbb{E}[Z_{n-1}|Z_{n-1} > 0]$, $k = q^{-1/4}$ and let

$$A = \{N \leq k, |D| \leq k, R_{n-1} > 2k\},$$

and

$$f(x) = (x - k)^{-1/2}I_{x \geq 2k+1}.$$

Using (4.4), (4.5), (4.6), and the fact that $\|f\| \leq k^{-1/2}$, we find

$$\mathbb{E}[f(X_q)] \leq q \sum_{j=1}^{\infty} \frac{(1-q)^j}{j^{1/2}} \leq (q\pi)^{1/2}$$

to obtain

$$\begin{aligned} \mathbb{E}[|D|S_1(W^e|\mathcal{F})\mathbf{I}_A] &\leq ku^{-1/2}\mathbb{E}f(R_{n-1}) \\ &\leq ku^{-1/2}(\mathbb{E}f(X_q) + qk^{-1/2}\mathbb{E}|D_{n-1}|) \\ &\leq Cq^{1/4}\log n, \end{aligned}$$

where $D_{n-1} = R_{n-1} - R_{n-1}^*$. Now, applying (2.4) yields

$$\mathbb{P}(R_{n-1} \leq 2k) \leq 1 - (1-q)^{2k} + q\mathbb{E}|D_{n-1}| \leq q(2k + \mathbb{E}|D_{n-1}|),$$

and by Markov's inequality and (4.4) we finally obtain

$$\mathbb{P}(A^c) \leq k^{-1}(\mathbb{E}N + \mathbb{E}|D|) + q(2k + \mathbb{E}|D|) \leq Cq^{1/4}\log n.$$

The theorem follows after using (v) and $\mathbb{E}Z_n = 1$ to obtain $\mathbb{E}[Z_n|Z_n > 0] \leq Cn$. \square

5 APPLICATION TO THE UNIFORM ATTACHMENT RANDOM GRAPH MODEL

Let G_n be a directed random graph on n nodes defined by the following recursive construction. Initially the graph starts with one node with a single loop where one end of the loop contributes to the “in-degree” and the other to the “out-degree.” Now, for $2 \leq m \leq n$, given the graph with $m-1$ nodes, add node m along with an edge directed from m to a node chosen uniformly at random among the m nodes present. Note that this model allows edges connecting a node with itself. This random graph model is referred to as uniform attachment.

This model has been well studied, and it was shown in Bollobas et al. (2001) that if W is equal to the in-degree of a node chosen uniformly at random from G_n , then W converges to a geometric distribution (starting at 0) with parameter $1/2$ as $n \rightarrow \infty$. We will give an explicit bound on the total variation distance between the distribution of W and the geometric distribution that yields this asymptotic. Some related results for this and other random graph models were given using Stein’s method in Ford (2009), where the author obtained the same rate found in our result below (but with larger constant) using an ad hoc analysis of the model. As we will see, our framework grants easy access to the result of Ford (2009) for this model. It’s worthwhile noting that this type of result is not obtainable using exponential approximation.

Theorem 5.1. *If W is the in-degree of a node chosen uniformly at random from the random graph G_n generated according to uniform attachment, then*

$$d_{\text{TV}}(\mathcal{L}(W), \text{Ge}^0(\frac{1}{2})) \leq \frac{1}{n}.$$

Proof of Theorem 5.1. Let X_i have a Bernoulli distribution, independent of all else, with parameter $\mu_i := (n-i+1)^{-1}$, and let N be an independent random variable that is uniform on the integers $1, 2, \dots, n$. If we imagine that node $n+1-N$ is the randomly selected node, then it’s easy to see that we can write $W := \sum_{i=1}^N X_i$.

Next, let us prove that $\sum_{i=1}^{N-1} X_i$ has the discrete equilibrium distribution w.r.t. W . First note that we have for bounded f and every m ,

$$\mu_m \mathbb{E} \Delta f \left(\sum_{i=1}^{m-1} X_i \right) = \mathbb{E} \left[f \left(\sum_{i=1}^m X_i \right) - f \left(\sum_{i=1}^{m-1} X_i \right) \right],$$

where we use

$$\mathbb{E} f(X_m) - f(0) = \mathbb{E} X_m \mathbb{E} \Delta f(0)$$

and thus the fact that we can write $(X_m)^{e_0} \equiv 0$. Note also that for any bounded function g with $g(0) = 0$ we have

$$\mathbb{E}\left(\frac{g(N)}{\mu_N} - \frac{g(N-1)}{\mu_N}\right) = \mathbb{E}g(N).$$

We now assume that $f(0) = 0$. Hence, using the above two facts and independence between N and the sequence X_1, X_2, \dots , we have

$$\mathbb{E}W\mathbb{E}\Delta f\left(\sum_{i=1}^{N-1} X_i\right) = \mathbb{E}f(W).$$

Now, let

$$N' = \begin{cases} N & \text{if } 1 \leq N < n, \\ 0 & \text{if } N = n. \end{cases}$$

We have that $\mathcal{L}(N') = \mathcal{L}(N-1)$ so that

$$W^{e_0} := \sum_{i=1}^{N'} X_i$$

has the equilibrium distribution with respect to W and it is plain that

$$\mathbb{P}[W \neq W^{e_0}] \leq \mathbb{P}[N = n] = \frac{1}{n}.$$

Applying (2.9) of Remark 2.5 yields the theorem. \square

6 APPLICATION TO THE PREFERENTIAL ATTACHMENT RANDOM GRAPH MODEL

Define the directed graph G_n on n nodes by the following recursive construction. Initially the graph starts with one node with a single loop where one end of the loop contributes to the “in-degree” and the other to the “out-degree.” Now, for $2 \leq m \leq n$, given the graph with $m-1$ nodes, add node m along with an edge directed from m to a random node chosen proportional to the total degree of the node. Note that at step m , the chance that node m connects to itself is $1/(2m-1)$ since we consider the added vertex m as immediately having out-degree equal to one. This random graph model is referred to as preferential attachment.

This model has been well studied, and it was shown in Bollobas et al. (2001) that if W is equal to the in-degree of a node chosen uniformly at

random from G_n , then W converges to the Yule-Simon distribution (defined below). We will give a rate of convergence in the total variation distance for this asymptotic, a result that cannot be read from the main results of Bollobas et al. (2001). Some rates of convergence in this and related random graph models can be found in the thesis Ford (2009), but the techniques and results there do not appear to overlap with ours below.

We say the random variable Z has the Yule-Simon distribution if

$$\mathbb{P}(Z = k) = \frac{4}{k(k+1)(k+2)}, \quad k = 1, 2, \dots$$

The following is our main result.

Theorem 6.1. *Let $W_{n,i}$ be the total degree of vertex i in the preferential attachment graph on n vertices and let I uniform on $\{1, \dots, n\}$ independent of $W_{n,i}$. If Z has the Yule-Simon distribution, then*

$$d_{\text{TV}}(\mathcal{L}(W_{n,I}), \mathcal{L}(Z)) \leq \frac{C \log n}{n}$$

for some constant C independent of n .

Remark 6.1. The notation $\mathcal{L}(W_{n,I})$ in the statement of Theorem 6.1 should be interpreted as

$$\mathcal{L}(W_{n,I} | I = i) = \mathcal{L}(W_{n,i}).$$

We will use similar notation without remark frequently in the sequel.

At this point, the reader may wonder where the geometric distribution enters into the discussion above. The following elementary proposition clarifies this point.

Proposition 6.2. *If U has the uniform distribution on $(0, 1)$, and given U , we define Z such that $\mathcal{L}(Z) = \text{Ge}(\sqrt{U})$, then Z has the Yule-Simon distribution.*

Our strategy to prove Theorem 6.1 will be to show that for I uniform on $\{1, \dots, n\}$ and U uniform on $(0, 1)$ we have

1. $d_{\text{TV}}(\mathcal{L}(W_{n,I}), \text{Ge}(\mathbb{E}[W_{n,I} | I]^{-1})) \leq C \log(n)/n,$
2. $d_{\text{TV}}(\text{Ge}(\mathbb{E}[W_{n,I} | I]^{-1}), \text{Ge}(\sqrt{I/n})) \leq C \log(n)/n,$
3. $d_{\text{TV}}(\text{Ge}(\sqrt{I/n}), \text{Ge}(\sqrt{U})) \leq C \log(n)/n,$

where here and in what follows we use the letter C as a generic constant which may differ from line to line. From this point, Theorem 6.1 follows from the triangle inequality and Proposition 6.2.

Item 1 will follow from our framework above; in particular we will use the following result which may be of independent interest. We postpone the proof to the end of the section.

Theorem 6.3. *Retaining the notation and definitions above, we have*

$$d_{\text{TV}}(\mathcal{L}(W_{n,i}), \text{Ge}(\mathbb{E}(W_{n,i})^{-1})) \leq \frac{C}{i}$$

for some constant C independent of n and i .

To show Items 2 and 3 we will need the following lemma. The first statement is found in Bollobas et al. (2001), p. 283 and the second follows easily from the first.

Lemma 6.4 (Bollobas et al. (2001)). *Retaining the notation and definitions above, for all $1 \leq i \leq n$,*

$$\left| \mathbb{E}W_{n,i} - \sqrt{\frac{n}{i}} \right| \leq C\sqrt{\frac{n}{i^3}} \quad \text{and} \quad \left| \frac{1}{\mathbb{E}W_{n,i}} - \sqrt{\frac{i}{n}} \right| \leq \frac{C}{\sqrt{ni}}.$$

Our final general lemma is useful for handling total variation distance for conditionally defined random variables.

Lemma 6.5. *Let W and V be random variables and let X be a random element defined on the same probability space. Then*

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(V)) \leq \mathbb{E} d_{\text{TV}}(\mathcal{L}(W|X), \mathcal{L}(V|X)).$$

Proof. If $f : \mathbb{R} \rightarrow [0, 1]$, then

$$\begin{aligned} |\mathbb{E}[f(W) - f(V)]| &\leq \mathbb{E}|\mathbb{E}[f(W) - f(V)|X]| \\ &\leq \mathbb{E} d_{\text{TV}}(\mathcal{L}(W|X), \mathcal{L}(V|X)). \end{aligned} \quad \square$$

Armed with these lemmas, we can prove Theorem 6.1.

Proof of Theorem 6.1. We first claim that

$$d_{\text{TV}}(\text{Ge}(p), \text{Ge}(p - \varepsilon)) \leq \frac{\varepsilon}{p} \leq \frac{\varepsilon}{p - \varepsilon}. \quad (6.1)$$

The second inequality of (6.1) is obvious and the first follows by coupling the Bernoulli sequences that generate the geometric variables via an infinite i.i.d. sequence of random variables uniform on $(0, 1)$ in the usual way.

Using (6.1) and Lemma 6.4 we easily obtain

$$d_{\text{TV}}(\text{Ge}(1/\mathbb{E}W_{n,i}), \text{Ge}(\sqrt{i/n})) \leq \frac{C}{i}$$

and applying Lemma 6.5 we find

$$d_{\text{TV}}(\text{Ge}(\mathbb{E}[W_{n,I}|I]^{-1}), \text{Ge}(\sqrt{I/n})) \leq \frac{C \log(n)}{n},$$

which is Item 2 above.

Now, Coupling U to I by writing $U = I/n - V$, where V is uniform on $(0, 1/n)$ and independent of I , and using first (6.1) and then Lemma 6.5 leads to

$$d_{\text{TV}}(\text{Ge}(\sqrt{U}), \text{Ge}(\sqrt{I/n})) \leq \frac{C}{n} \sum_{i=1}^n \frac{\sqrt{i/n} - \sqrt{(i-1)/n}}{\sqrt{i/n}} \leq \frac{C \log(n)}{n},$$

which is Item 3 above.

Finally, applying Lemma 6.5 to Theorem 6.3 yields the claim related to Item 1 above so that Theorem 6.1 is proved. \square

The remainder of this section is devoted to the proof of Theorem 6.3; recall $W_{n,i}$ is the total degree of vertex i in the preferential attachment graph on n vertices. Since we want to apply our geometric approximation framework using the equilibrium distribution, we will use Proposition 2.3 and so we first construct a variable having the size-bias distribution of $W_{n,i} - 1$. To facilitate this construction we need some auxiliary variables.

For $j \geq i$, let $X_{j,i}$ be the indicator variable of the event that vertex j has an outgoing edge connected to vertex i in G_j so that we can denote $W_{j,i} = 1 + \sum_{k=i}^j X_{k,i}$. In this notation, for $1 \leq i < j \leq n$,

$$\mathbb{P}(X_{j,i} = 1 | G_{j-1}) = \frac{W_{j-1,i}}{2j-1},$$

and for $1 \leq i \leq n$,

$$\mathbb{P}(X_{i,i} = 1 | G_{i-1}) = \frac{1}{2i-1}.$$

The following well known result (see Proposition 2.2 and the discussion after in Chen, Goldstein, and Shao (2011)) will allow us to use this decomposition to size-bias $W_{n,i} - 1$.

Proposition 6.6. *Chen et al. (2011) Let X_1, \dots, X_n be zero-one random variables with $\mathbb{P}(X_i = 1) = p_i$. For each $i = 1, \dots, n$, let $(X_j^{(i)})_{j \neq i}$ have the distribution of $(X_j)_{j \neq i}$ conditional on $X_i = 1$. If $X = \sum_{i=1}^n X_i$, $\mu = \mathbb{E}[X]$, and K is chosen independent of the variables above with $\mathbb{P}(K = k) = p_k/\mu$, then $X^s = \sum_{j \neq K} X_j^{(K)} + 1$ has the size-bias distribution of X .*

Roughly, Proposition 6.6 implies that in order to size-bias $W_{n,i} - 1$, we choose an indicator $X_{K,i}$ where for $k = i, \dots, n$, $\mathbb{P}(K = k)$ is proportional to $\mathbb{P}(X_{n,k} = 1)$ (and zero for other values), then attach vertex K to vertex i and sample the remaining edges conditional on this event. Note that given $K = k$, in the graphs G_l , $1 \leq l < i$ and $k < l \leq n$, this conditioning does not change the original rule for generating the preferential attachment graph given G_{l-1} . The following lemma implies the remarkable fact that in order to generate the graphs G_l for $i \leq l < k$ conditional on $X_{k,i} = 1$ and G_{l-1} , we attach edges following the same rule as preferential attachment, but include the edge from vertex k to vertex i in the degree count.

Lemma 6.7. *Retaining the notation and definitions above, for $i \leq j < k$ we have*

$$\mathbb{P}(X_{j,i} = 1 | X_{k,i} = 1, G_{j-1}) = \frac{1 + W_{j-1,i}}{2j},$$

where we define $W_{i-1,i} = 1$.

Proof. By Bayes' rule, we have

$$\begin{aligned} & \mathbb{P}(X_{j,i} = 1 | X_{k,i} = 1, G_{j-1}) \\ &= \frac{\mathbb{P}(X_{j,i} = 1 | G_{j-1}) \mathbb{P}(X_{k,i} = 1 | X_{j,i} = 1, G_{j-1})}{\mathbb{P}(X_{k,i} = 1 | G_{j-1})}, \end{aligned} \quad (6.2)$$

and we will calculate the three probabilities appearing in (6.2). First, for $i \leq j$, we have

$$\mathbb{P}(X_{j,i} = 1 | G_{j-1}) = \frac{W_{j-1,i}}{2j-1}.$$

which implies

$$\mathbb{P}(X_{k,i} = 1 | G_{j-1}) = \frac{\mathbb{E}[W_{k-1,i} | G_{j-1}]}{2k-1}$$

and

$$\mathbb{P}(X_{k,i} = 1 | X_{j,i} = 1, G_{j-1}) = \frac{\mathbb{E}[W_{k-1,i} | X_{j,i} = 1, G_{j-1}]}{2k-1}.$$

Now, to compute the conditional expectations appearing above, note first that

$$\mathbb{E}(W_{k,i}|G_{k-1}) = W_{k-1,i} + \frac{W_{k-1,i}}{2k-1} = \left(\frac{2k}{2k-1}\right) W_{k-1,i},$$

and thus

$$\mathbb{E}(W_{k,i}|G_{k-2}) = \left(\frac{2(k-1)}{2(k-1)-1}\right) \left(\frac{2k}{2k-1}\right) W_{k-2,i}.$$

Iterating, we find that for $i, s < k$,

$$\mathbb{E}(W_{k,i}|G_{k-s}) = \prod_{t=1}^s \left(\frac{2(k-t+1)}{2(k-t+1)-1}\right) W_{k-s,i}. \quad (6.3)$$

By setting $j-1 = k-s$ and then replacing $k-1$ by k in (6.3) we obtain

$$\mathbb{E}(W_{k-1,i}|G_{j-1}) = \prod_{t=1}^{k-j} \left(\frac{2(k-t)}{2(k-t)-1}\right) W_{j-1,i}$$

which also implies

$$\mathbb{E}(W_{k-1,i}|X_{j,i} = 1, G_{j-1}) = \prod_{t=1}^{k-j-1} \left(\frac{2(k-t)}{2(k-t)-1}\right) (1 + W_{j-1,i}).$$

Substituting these expressions appropriately into (6.2) proves the lemma. \square

The previous lemma suggests the following (embellished) construction of $(W_{n,i}|X_k = 1)$ for any $1 \leq i \leq k \leq n$. Here and below we will denote quantities related to this construction with a superscript k . First we generate G_{i-1}^k , a graph with $i-1$ vertices, according to the usual preferential attachment model. At this point, if $i \neq k$, vertex i and k are added to the graph, along with a vertex labelled i' with an edge to it emanating from vertex k . Given G_{i-1}^k and these additional vertices and edges, we generate G_i^k by connecting vertex i to a vertex j randomly chosen from the vertices $1, \dots, i, i'$ proportional to their degree, where i has degree one (from the out-edge) and i' has degree one (from the in-edge emanating from vertex k). If $i = k$, we attach i to i' and denote the resulting graph by G_i^i . For $i < j < k$, we generate the graphs G_j^k recursively from G_{j-1}^k by connecting vertex j to a vertex l randomly chosen from the vertices $1, \dots, j, i'$ proportional to their

degree, where j has degree one (from the out-edge). Note that none of the vertices $1, \dots, k-1$ can connect to vertex k . We now define $G_k^k = G_{k-1}^k$, and for $j = k+1, \dots, n$, we generate G_j from G_{j-1} according to preferential attachment among the vertices $1, \dots, j, i'$.

The following lemma summarizes relevant properties of this construction.

Lemma 6.8. *Let $1 \leq i \leq k \leq n$ and retain the notation and definitions above.*

1. $\mathcal{L}(W_{n,i}^k + W_{n,i'}^k) = \mathcal{L}(W_{n,i}|X_k = 1)$.
2. For fixed i , if K is a random variable such that

$$\mathbb{P}(K = k) = \frac{\mathbb{E}X_{k,i}}{\mathbb{E}W_{n,i} - 1}, \quad k \geq i,$$

then $W_{n,i}^K + W_{n,i'}^K - 1$ has the size-bias distribution of $W_{n,i} - 1$.

3. Conditional on the event $\{W_{n,i}^k + W_{n,i'}^k = m + 1\}$, the variable $W_{n,i}^k$ is uniformly distributed on the integers $1, 2, \dots, m$.
4. $W_{n,i}^K - 1$ has the discrete equilibrium distribution of $W_{n,i} - 1$.

Proof. Items 1 and 2 follow from Proposition 6.6 and Lemma 6.7. Viewing $(W_{n,i}^k, W_{n,i'}^k)$ as the number of balls of two colors in a Polya urn model started with one ball of each color, Item 3 follows from induction on m , and Item 4 follows from Proposition 2.3. \square

With this setup, we are ready to give a proof of Theorem 6.3.

Proof of Theorem 6.3. We will apply Theorem 2.2 to $\mathcal{L}(W_{n,i} - 1)$, so that we must find a coupling of a variable with this distribution to that of a variable having its discrete equilibrium distribution.

For each fixed $k = i, \dots, n$ we will construct $(X_{j,i}^k, \tilde{X}_{j,i}^k)_{j \geq i}$ so $(X_{j,i}^k)_{j \geq i}$ and $(\tilde{X}_{j,i}^k)_{j \geq i}$ are distributed as the indicators of the events vertex j connects to vertex i in G_n^k and G_n , respectively. We will use the notation

$$W_{j,i}^k = \sum_{m=i}^j X_{m,i}^k \quad \text{and} \quad \tilde{W}_{j,i}^k = \sum_{m=i}^j \tilde{X}_{m,i}^k,$$

which will be distributed as the degree of vertex i in the appropriate graphs.

The constructions for $k = i$ and $k > i$ differ, so assume here that $k > i$. Let $U_{j,i}^k$ be independent uniform $(0, 1)$ random variables and first define

$$X_{i,i}^k = \mathbb{I}[U_{i,i}^k < 1/2i] \quad \text{and} \quad \tilde{X}_{i,i}^k = \mathbb{I}[U_{i,i}^k < 1/(2i-1)].$$

Now, for $i < j < k$, and assuming that $(W_{j-1,i}^k, \widetilde{W}_{j-1,i}^k)$ is given, we define

$$X_{j,i}^k = \mathbb{I}\left[U_{j,i}^k < \frac{W_{j-1,i}^k}{2j}\right] \quad \text{and} \quad \tilde{X}_{j,i}^k = \mathbb{I}\left[U_{j,i}^k < \frac{\widetilde{W}_{j-1,i}^k}{2j-1}\right]. \quad (6.4)$$

For $j = k$ we set $X_{k,i}^k = 0$ and $\tilde{X}_{k,i}^k$ as in (6.4) with $j = k$, and for $j > k$ we define

$$X_{j,i}^k = \mathbb{I}\left[U_{j,i}^k < \frac{W_{j-1,i}^k}{2j-1}\right] \quad \text{and} \quad \tilde{X}_{j,i}^k = \mathbb{I}\left[U_{j,i}^k < \frac{\widetilde{W}_{j-1,i}^k}{2j-1}\right]$$

Thus we have recursively defined the variables $(X_{j,i}^k, \tilde{X}_{j,i}^k)$ and it is clear they are distributed as claimed with $(W_{j,i}^k, \widetilde{W}_{j,i}^k)$ distributed as the required degree counts. Note also that $\widetilde{W}_{j,i}^k \geq W_{j,i}^k$ and $\tilde{X}_{j,i}^k \geq X_{j,i}^k$. We also define the events

$$A_{j,i}^k := \{\min\{i \leq l \leq n : X_{l,i}^k \neq \tilde{X}_{l,i}^k\} = j\}.$$

Using that $W_{j-1,i}^k = \widetilde{W}_{j-1,i}^k$ under $A_{j,i}^k$ (which also implies $A_{j,i}^k = \emptyset$ for $j > k$) we have

$$\begin{aligned} \mathbb{P}(\widetilde{W}_{n,i}^k \neq W_{n,i}^k) &= \mathbb{P}\left(\bigcup_{j=i}^n A_{j,i}^k\right) \\ &= \mathbb{E}\tilde{X}_{k,i}^k + \sum_{j=i}^k \mathbb{P}\left(A_{j,i}^k \cap \left\{\frac{W_{j-1,i}^k}{2j} < U_{j,i}^k < \frac{\widetilde{W}_{j-1,i}^k}{2j-1}\right\}\right) \\ &\leq \mathbb{E}\tilde{X}_{k,i}^k + \sum_{j=i}^n \mathbb{P}\left(\frac{\widetilde{W}_{j-1,i}^k}{2j} < U_{j,i}^k < \frac{\widetilde{W}_{j-1,i}^k}{2j-1}\right), \end{aligned} \quad (6.5)$$

where we write $W_{i-1,i}^k := \widetilde{W}_{i-1,i}^k := 1$. Finally, starting from (6.5) and using the computations in the proof of Lemma 6.7 and the estimates in Lemma 6.4, we find

$$\mathbb{P}(\widetilde{W}_{n,i}^k \neq W_{n,i}^k) \leq \frac{\mathbb{E}W_{k-1,i}^k}{2k-1} + \sum_{j=i}^n \mathbb{E}W_{j-1,i}^k \left(\frac{1}{2j-1} - \frac{1}{2j}\right)$$

$$\begin{aligned} &\leq C \left[\sqrt{\frac{k}{i}} \frac{1}{k} + \sqrt{\frac{k}{i^3}} \frac{1}{k} + \sum_{j \geq i} \left(\sqrt{\frac{j}{i}} \frac{1}{j^2} + \sqrt{\frac{j}{i^3}} \frac{1}{j^2} \right) \right] \\ &\leq C/i. \end{aligned}$$

If $k = i$, it is clear from the construction preceding Lemma 6.8 that an easy coupling similar to that above will yield $\mathbb{P}(\widetilde{W}_{n,i}^i \neq W_{n,i}^i) < C/i$. Since these bounds do not depend on k , we also have

$$\mathbb{P}(\widetilde{W}_{n,i}^K - 1 \neq W_{n,i}^K - 1) \leq C/i, \quad (6.6)$$

and the result now follows from Lemma 6.8, (6.6), and (2.9) of Remark 2.5. \square

REFERENCES

- D. J. Aldous (1989). *Probability approximations via the Poisson clumping heuristic*, volume 77 of *Applied Mathematical Sciences*. Springer-Verlag, New York. ISBN 0-387-96899-7.
- A. D. Barbour and V. Čekanavičius (2002). Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30**, 509–545.
- A. D. Barbour and R. Grübel (1995). The first divisible sum. *J. Theoret. Probab.* **8**, 39–47.
- B. Bollobás, O. Riordan, J. Spencer, G. Tusnády (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* 18, no. 3, 279–290.
- M. Brown (1990). Error bounds for exponential approximations of geometric convolutions. *Ann. Probab.* **18**, 1388–1402.
- L.H.Y. Chen, L. Goldstein, and Q.M. Shao (2011). *Normal approximation by Stein's method*. Springer, Heidelberg. ISBN 978-3-642-15006-7.
- F. Daly (2010). Stein's method for compound geometric approximation. *Journal of Applied Probability* **47**, 146–156.
- F. Daly (2008). Upper bounds for Stein-type operators. *Electronic Journal of Probability* **13**, 566–587.
- E. Ford (2009) Barabási-Albert random graphs, scale-free distributions and bounds for approximation through Steins method. *Ph.D. thesis, Jesus College, University of Oxford*.
- L. Goldstein (2009). Personal communication and unpublished notes. Stein workshop, January 2009, Singapore.
- L. Goldstein and G. Reinert (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935–952.
- V. Kalashnikov (1997) *Geometric sums: bounds for rare events with applications*. Kluwer Academic Publishers Group, Dordrecht

- S. P. Lalley and X. Zheng (in press). Occupation statistics of critical branching random walks in two or higher dimensions. *Ann. Probab.*
- R. Lyons, R. Pemantle, and Y. Peres (1995). Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23**, 1125–1138.
- L. Mattner and B. Roos (2007). A shorter proof of Kanter’s Bessel function concentration bound. *Probab. Theory Related Fields* **139**, 191–205.
- E. A. Peköz (1996). Stein’s method for geometric approximation. *J. Appl. Probab.* **33**, 707–713.
- E. Peköz and A. Röllin (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*, Vol. 39, No. 2, 587608.
- M. J. Phillips and Graham V. Weinberg (2000). Non-uniform bounds for geometric approximation. *Statist. Probab. Lett.* **49**, 305–311.
- A. Röllin (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.
- A. Röllin (2008). Symmetric binomial approximation for sums of locally dependent random variables. *Electron. J. Probab.* **13**, 756–776.
- S. Ross and E. Peköz (2007). *A second course in probability*. www.ProbabilityBookstore.com, Boston, MA.
- A. M. Yaglom (1947). Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)* **56**, 795–798.